PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 1/04

A2

(11) International Publication Number: **WO 97/08190**

(43) International Publication Date: 6 March 1997 (06.03.97)

(21) International Application Number: PCT/EP96/03731

(22) International Filing Date: 23 August 1996 (23.08.96)

(30) Priority Data:
9517661.6 30 August 1995 (30.08.95) GB

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: HUGHES, Ian [GB/GB]; SmithKline Beecham Pharmaceuticals, New Frontiers Science Park South, Third Avenue, Harlow, Essex CM19 5AW (GB).

(74) Agent: SUMMERSELL, Richard, John; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: COMPOUNDS

(57) Abstract

The invention relates to combinatorial chemistry, in particular synthesis of combinatorial libraries.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

COMPOUNDS

The present invention relates to combinatorial chemistry, in particular the synthesis of combinatorial libraries which can be used for the identification of bioactive molecules.

In recent years, there has been a growing demand for the production and identification of small molecules that have pharmacological activity as, for example, agonists or antagonists of various cellular acceptor molecules, such as cell-surface receptors, enzymes or antibodies. Historically, the process to identify pharmacologically active molecules has been characterised by the preparation of single compounds followed by biological testing, which is time-consuming, labour intensive and inefficient.

In an effort to reduce the time and expense involved in screening a large number of randomly chosen compounds for biological activity, several developments have been made in using peptides or nucleotides combined with solid phase synthesis to provide libraries of compounds for the discovery of lead compounds. See, for example, Lebl et al., *Int. J. Pept. Prot. Res.*, **41**, p. 201 (1993) which discloses methodologies providing selectively cleavable linkers between peptide and resin such that a certain amount of peptide can be liberated from the resin and assayed in soluble form while some of the peptide still remains attached to the resin, where it can be sequenced; Lam et al., *Nature*, **354**, p. 82 (1991) and WO 92/00091 which disclose a method of synthesis of linear peptides on a solid support such as polystyrene or polyacrylamide resin; Geysen et al., *J. Immunol. Meth.*, **102**, p. 259 (1987) which discloses the synthesis of peptides on derivatized polystyrene pins which are arranged on a block in such a way that they correspond to the arrangement of wells in a 96-well microtiter plate; and Houghten et al., *Nature*, **354**, p. 84 (1991) and WO 92/09300 which disclose an approach to *de novo* determination of antibody or receptor binding sequences involving soluble peptide pools.

By the term "library" or "combinatorial library" is meant a collection of individual compounds, each compound having a common core structure wherein the library contains a discrete number of independently variable substituents, functional groups or structural elements, and further, wherein the library is designed so that, for the range of chemical moieties selected for each of the independently variable substituents, compounds containing all possible permutations of those substituents will be present in the library. Thus, if a core structure, labelled R, contains three independently variable substituents, labelled X, Y and Z, and if X is taken from *m* different chemical moieties, Y from *n* different chemical moieties and Z from *p*

different chemical moieties (wherein *m*, *n* and *p* are integers which define the size of the library, and which range between 1 to 1000; preferably between 1 to 100; most preferably between 1 to 20), then the library would contain *m* x *n* x *p* different chemical compounds and all possible combinations of X, Y and Z would be present on the core structure R within that library. A typical library will typically contain between 2 to 10000 or more compounds, and often more than 10000 compounds.

Once the library of compounds has been synthesised and screened there must be some method to deconvolute the results of screening such that individual active compounds can be identified. Many methods of deconvolution are possible. For example the library can be deconvoluted by an iterative approach, which involves the re-synthesis of mixtures of decreasing complexity until a single compound is identified. In particular, sub-libraries can themselves be screened. For example if a main library of 100 components is active, 10 sub-libraries of 10 components can be screened. This approach is advantageous since in a sub-library one of the substituents, i.e. the last substituent to be introduced, can be defined and kept constant. However the approach has the major disadvantage of being time consuming.

After deconvolution of the library, a single compound, or relatively small number of compounds, are usually identified which have the desired biological activity. This compound can then serve as a lead for the preparation of further structurally related libraries or single compounds. In the case where libraries are synthesised on beads, a collection of beads are usually screened, and if biological activity is detected the single beads are screened and the active compound identified.

It will be apparent that the identification of a biologically active compound from a complex mixture is time consuming and presents certain difficulties. Therefore there is a need for improved methods for efficiently identifying pharmaceutically active compounds prepared as libraries.

One approach to this problem is the concept of "tagging" synthesis beads by the introduction of chemical "tags" at each synthetic step during the construction of the library. The nature of the chemical tags on a particular synthesis bead thus defines the synthetic history of that bead and facilitates structure determination. See, for example, Kerr et al., *J. Amer. Chem. Soc.*, **115**, p. 2529 (1993) which discloses the use of peptide coding strands, which can be "read" by Edman sequencing, and Ohlmeyer et al., *Proc . Natl. Acad. Sci.*, **90**, p. 10922 (1993) which discloses molecular tags which can be "read" by electron capture capillary gas chromatography. A key requirement for compound identification using tagged beads

is that the compound must remain bound to or associated with its bead of origin throughout the screening procedure.

Certain screening strategies may rely on the separation of biologically active compounds from soluble mixtures. For example, Chu et al., *J. Amer. Chem. Soc.* **117**, p.5419 (1995) disclose the use of affinity capillary electrophoresis for the separation from combinatorial libraries of those ligands that bind most tightly to a receptor. In such screening strategies any former association of a compound with its bead of origin, and hence structural information contained in "tags", would be lost.

It is also established that mass spectroscopy can be used to identify compounds from single synthesis beads. For example, Chen et al., *J. Amer. Chem. Soc.* **116**, p. 2661 (1994), and Stankova et al., *Drug Dev. Res.*, **33**,p.146 (1994). However, when large numbers of compounds are present in a combinatorial library it is highly likely that many will have the same nominal molecular weights. This leads to a phenomenon known as "mass-redundancy" by which is meant that compounds are indistinguishable on the basis of molecular weight alone. Mass-redundancy may be reduced by measuring molecular weights at higher resolution, when, ultimately, only compounds having identical empirical formulae would be indistinguishable.

The present invention is based on the principle that each compound in a library will have, by design, a unique molecular weight which can serve as an identifier for that particular compound. The invention provides a method for the identification of a biologically active compound, and in particular, the identification of a compound derived from a single biologically active bead. The advantages of this method over tagging synthesis beads are firstly the present invention does not impose any restrictions on the nature of the chemistry used to synthesise the combinatorial library, since it does not have to be compatible with tagging chemistry and does not introduce additional non-productive synthetic steps, and secondly by using the present invention the compound can, if so required, be identified without association with the bead of origin. The fact that tagging is not required is clearly advantageous. An additional advantage is the ability to identify the compound by its nominal mass without recourse to high resolution mass spectrometry. Of course, although not necessary for identification, high resolution measurements and analysis of fragmentation patterns remain available options for further confimatory evidence.

The present invention provides a method for the control of mass redundancies in a combinatorially synthesised compound library which comprises identifying compounds by their molecular weight

Preferably molecular weight is determined by mass spectrometry. Preferably the above method is used to identify compounds derived from a single bead.

In situations where two or more compounds have identical nominal molecular mass, the present invention allows for the deliberate incorporation of the natural isotopic mass patterns of chorine and bromine atoms or other artificially isotopically enriched atoms or molecules, to further extend its scope and usefulness. It will be apparent to those skilled in the art that a structure designed by the method of this invention can be unambiguously characterised by determining its nominal mass and isotope pattern.

The invention will now be explained and exemplified in detail.

1. Background to the invention

The present invention relies on a selection strategy which is based on the following observation. When consecutive integers are arrayed in a table of c columns and r rows in the manner exemplified below, and two sets of numbers are selected such that Set 1 comprises numbers all from the same column (maximum set size r), and Set 2 comprises numbers each selected from different columns (maximum set size c), the addition of any pair of numbers, one chosen from each set, will generate a sum which differs from that created by any other similar pairwise combination, giving a maximum of (r x c) unique combinations.

Example 1.1 Array of 100 integers (c=10 and r=10)

1	2	*3*	4	5	6	7	8	9	10
11	12	*13*	14	15	16	**17**	18	19	20
21	22	*23*	24	25	26	27	**28**	29	30
31	**32**	*33*	34	35	36	37	38	39	40
41	42	*43*	44	45	46	47	48	49	50
51	52	*53*	54	55	56	57	58	59	60
61	62	*63*	**64**	65	66	67	68	69	70
71	72	*73*	74	75	76	77	78	79	80
81	82	*83*	84	85	**86**	87	88	89	**90**
91	92	*93*	94	**95**	96	97	98	99	100

The largest sets that can be selected from this table each contain 10 numbers, and the pairwise addition of two such sets, Set 1 (same column) in *italic*, Set 2 (different columns) in **bold** is shown below. It is clear by inspection that all sums are unique.

For this particular selection, the 100 unique values fall in the range 12 to 188. It is also apparent that the same number (in this case *23*) may occur in both sets.

Set1+Set2	**31**	**32**	**23**	**64**	**95**	**86**	**17**	**28**	**9**	**90**
3	34	35	26	67	98	89	20	31	12	93
13	44	45	36	77	108	99	30	41	22	103
23	54	55	46	87	118	109	40	51	32	113
33	64	65	56	97	128	119	50	61	42	123
43	74	75	66	107	138	129	60	71	52	133
53	84	85	76	117	148	139	70	81	62	143
63	94	95	86	127	158	149	80	91	72	153
73	104	105	96	137	168	159	90	101	82	163
83	114	115	106	147	178	169	100	111	92	173
93	124	125	116	157	188	179	110	121	102	183

Example 1.2 Array of 78 integers (c=13 and r=6)

1	**2**	*3*	4	5	6	7	**8**	9	10	**11**	12	13
14	15	*16*	17	18	19	20	21	22	23	24	25	26
27	28	*29*	30	**31**	32	33	34	35	36	37	38	39
40	41	*42*	43	44	**45**	**46**	47	48	**49**	50	51	**52**
53	54	*55*	**56**	57	58	59	60	61	62	63	64	65
66	67	*68*	69	70	71	72	73	**74**	75	76	**77**	78

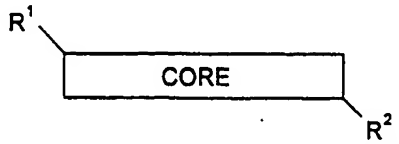
This table illustrates the effect of altering the number of columns (c), namely that as c varies, different integers are brought into alignment within a given column. The significance of this observation will be explained later.

The pairwise addition of two sets chosen from this 13 x 6 table is shown below and gives rise to 78 unique values in the range 5 to 145.

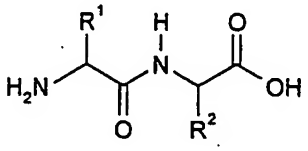
Set1+Set2	**14**	**2**	**3**	**56**	**31**	**45**	**46**	**8**	**74**	**49**	**11**	**77**	**52**
3	17	5	6	59	34	48	49	11	77	52	14	80	55
16	30	18	19	72	47	61	62	24	90	65	27	93	68
29	43	31	32	85	60	74	75	37	103	78	40	106	81
42	56	44	45	98	73	87	88	50	116	91	53	119	94
55	69	57	58	111	86	100	101	63	129	104	66	132	107
68	82	70	71	123	99	113	114	76	142	117	79	145	120

2. Application of method to selection of combinatorial library substituents

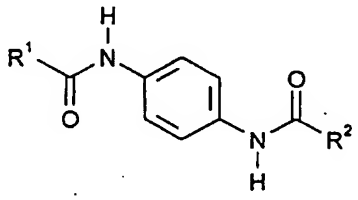
Combinatorial libraries may be defined as mixtures of related compounds having a common "core" structure which bears substituent groups (or R-groups) at a number of positions.



Within an individual library, all components have the same core structure, but have different substituent groups at one or more positions. Thus in a library of dipeptides, the common core structure might be the peptide backbone, and the varying R-groups would represent the amino acid side-chains.



In a typical non-peptide library the common core structure might be a multiply substituted ring system, for example a para-dianilide, with varying R-groups possibly derived from a range of different acylating agents.



The above method can be applied to substituent group selection by mapping the nominal molecular weights of available substituent groups onto tables of varying numbers of columns. The charts so generated are defined as having a "periodicity" equal to the number of columns. Different charts may be envisaged for different reagent types (e.g. R-COCl, R-NCO). The selection of sets of substituent groups must follow the following three rules.

Rule 1. Set 1 and Set 2 should be chosen from charts of the same periodicity.
Rule 2. Members of Set 1 should all be chosen from the **same** column (and be of different masses).
Rule 3. Members of Set 2 should all be chosen from **different** columns.

It is convenient, although by no means necessary, to map only the varying R-group of the reagents, i.e. the R of R-COCl, R-NCO, $HO_2CH(R)NH_2$, since the remainder of the group adds a constant mass and may be regarded as part of the core structure of the library. In the case of groups containing Cl and Br, only the lowest isotopic weight (35 or 79 respectively) is used.

Example 2.1 A small data set: the natural amino acids

The amino acids make a convenient sized data set to illustrate features of the method. The 19 amino acid side chains (excluding proline) denoted by the standard three-letter code of their parent amino acid, are mapped to their masses in Chart 1.

Chart 1. Amino acid side chains, periodicity 10

1 Gly	2	3	4	5	6	7	8	9	10
11	12	13	14	15 Ala	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31 Ser	32	33	34	35	36	37	38	39	40
41	42	43 Val	44	45 Thr	46	47 Cys	48	49	50
51	52	53	54	55	56	57 Leu Ile	58 Asn	59 Asp	60
61	62	63	64	65	66	67	68	69	70
71	72 Gln Lys	73 Glu	74	75 Met	76	77	78	79	80
81 His	82	83	84	85	86	87	88	89	90
91 Phe	92	93	94	95	96	97	98	99	100 Arg
101	102	103	104	105	106	107 Tyr	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130 Trp

Because the rules involve selection from columns, the chart produced can be much simplified by ignoring the mass numbers and condensing the columns. Isobaric groups are shown on the same line (Gln/Lys and Leu/Ile) - clearly only one from each pair may be chosen.

Chart 2. Amino acid side chains, periodicity 10

Gly Ser His Phe	Gln/Lys	Val Glu		Ala Thr Met		Cys Leu/Ile Tyr	Asn	Asp	Arg Trp

Set 1 (same column) may contain up to 4 groups,(e.g. Gly, Ser, His, Phe) and Set 2 (different columns) up to 8 (e.g. Gly, Lys, Val, Met, Leu, Asn, Asp, Trp). Hence a total of 32 unique dipeptides could be generated. If the excercise is repeated with a 16 column table, Chart 3 is produced (after simplification).

Chart 3. Amino acid side chains, periodicity 16

Gly His	Trp		Arg				Gln/Lys	Leu/Ile Glu	Asn	Val Asp Met Phe Tyr		Thr		Ala Ser Cys

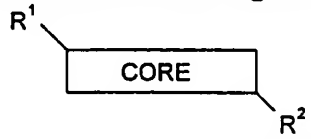
In this case up to 5 (Set1) x 9 (Set 2) = 45 unique nominal mass dipeptides could be generated.

Some general points can be observed from the above charts:

- The choice of substituent group is most limited for Set 1 chosen from the same column. In the case of Chart 2, there is only one way to select four substituents in a single column, i.e. Gly, Ser, His, Phe (Column 1). If only three groups were required, then 7 permutations would be possible, e.g. Gly, Ser, Phe (Column 1) or Cys, Leu, Tyr (Column 7).
- The number of permutations for choosing from different columns is considerable, in the case of Chart 2 it is 4x2x2x3x4x1x1x2 = 384.
- Changing the initial number of columns brings quite different sets of substituents into alignment. (e.g. Ala, Thr, Met in Chart 2 and Ala, Ser, Cys in Chart 3)

Example 2.2 Application of a larger data set to non-peptide libraries

Non-peptide libraries are generally non-oligomeric and often consist of a core structure bearing a number of substituent groups. Typically the substituent groups are derived from sets of similar reagents. For example, one or both of the groups R^1 and R^2 might be derived from the acid chloride reagents R^1COCl and R^2COCl.



A large number of acid chlorides are known in the chemical literature and many are available commercially. Chart 4 shows the R-groups of a small selection of acid chlorides mapped to their nominal molecular weights on a periodicity 10 chart, according to the method of this invention. The groups are labelled with their nominal molecular weight suffixed with a letter to distinguish isobaric groups. The use of Chart 4 to select sets of substituents for use in combinatorial libraries is exemplified below.

Selection of Set 1 using Rule 2 (selection from the same column).

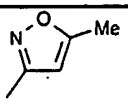
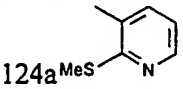
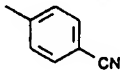
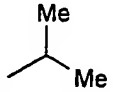
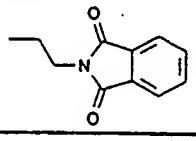
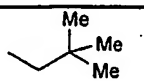
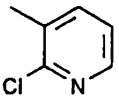
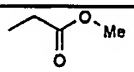
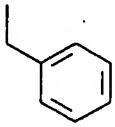
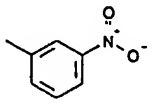
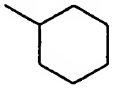
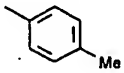
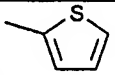
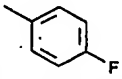
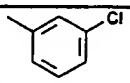
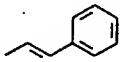
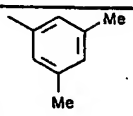
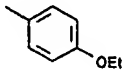
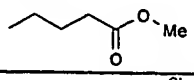
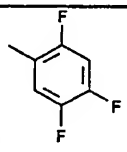
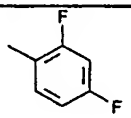
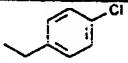
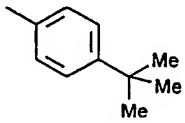
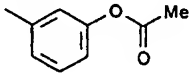
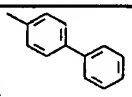
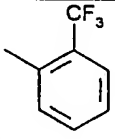
Inspection of Chart 4 shows that up to ten groups of different nominal mass may be selected from column 7, for example, the groups labelled 27a, 57a, 67a, 77a, 87a, 107a, 117a, 127a, 137a, and 147a. Numerous smaller sets could be selected from this and other columns of Chart 4.

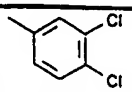
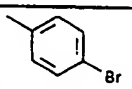
Selection of Set 2 using Rule 3 (selection from different columns)

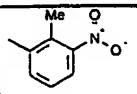
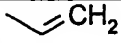
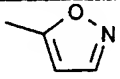
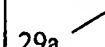
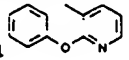
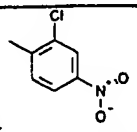
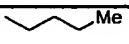
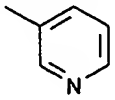
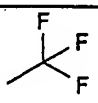
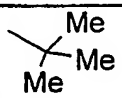
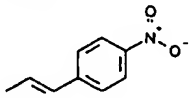
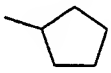
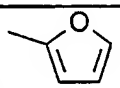
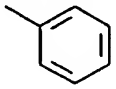
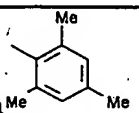
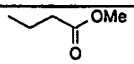
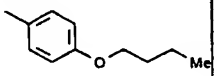
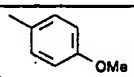
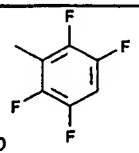
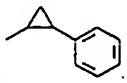
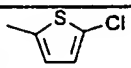
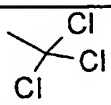
A maximum of ten groups may be selected from different columns of Chart 4, since this is the number of columns in the chart (periodicity = 10). Many such sets could be selected, for example a set comprising the groups labelled 71a, 102a, 83a, 174a, 145a, 136a, 127a, 78a, 69a and 170a.

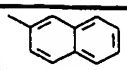
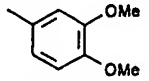
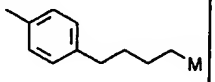
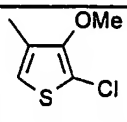
In combination with a suitable core structure Sets 1 and 2 above would give $10 \times 10 = 100$ compounds of different nominal mass.

Chart 4. Selection of acid chloride substituent groups, periodicity 10

Column 1	Column 2	Column 3	Column 4	Column 5
41a 	82a 	43a 	124a 	15a
71a 	102a 	43b 	174a 	45a
71b 	112a 	73a 		55a 
91a 	122a 	83a 		85a 
91b 		83b 		95a 
111a 		103a 		105a 
121a 		113a —$(CH_2)_7$-Me		115a 
131a 		113b 		125a 
		133a 		135a 
		153a 		145a 

				 145b
				 155a

Column 6	Column 7	Column 8	Column 9	Column 10
 136a	 27a	 68a	 29a	 170a
 156a	 57a	 78a	 69a	
	 57b	148a 	 69b	
	 67a		99a 	
	 77a		 119a	
	 87a		149a 	
	 107a		 149b	
	 117a			
	 117b			
	 117c			

	127a 			
	127b 			
	137a 			
	147a 			
	 147b			

3. Use of isotope patterns to extend scope of the method.

If two compounds have the same nominal mass (M), but contain different numbers of Cl and Br atoms, they may be distinguished by their M+2, M+4, etc. isotope patterns. Hence the selection rules may be extended as follows.

Rule 4. Either, but not both, of Rules 2 and 3 may be replaced by Rules 5 and 6 respectively.

Rule 5. Members of Set 1 should be chosen from the same column and may have the same mass provided that the isotope pattern (ie. Cl, Br count) is different.

Rule 6. Members of Set 2 should be chosen from any columns provided that all members chosen from an individual column have different isotope patterns (ie. different Cl, Br counts).

Example 3.1 Application of Rule 5 to extend Set 1

Inspection of column 7 of Chart 4 shows that all three groups at mass 117, i.e. groups (117a, 117b and 117c) may now be included since they contain 0, 1 and 3 chlorine atoms respectively. Similarly groups 147a and 147b with 0 and 1 chlorine atoms respectively may be included. In contrast, groups 127a and 127b may not both be included since they are isobaric and cannot be distinguished by their isotope content. The modified set, Set 1a contains 13 groups, namely 27a, 57a, 67a, 77a, 87a, 107a, 117a, 117b, 117c, 127a, 137a, 147a and 147b. In combination with Set 2, this modified set would give rise to 13 x 10 = 130 distinguishable compounds.

Example 3.2 Application of Rule 6 to extend Set 2

Examination of column 1 of Chart 4 reveals that group 111a may be used in addition to the previously selected group 71a since the groups contain 1 and 0 chlorine atoms respectively. Similarly in column 5, groups 125a (1 chlorine), 145b (2 chlorines) and 155a (1 bromine) may be added, as can groups 117b (1 chlorine) and 117c (3 chlorines) from column 7. The modified set, Set 2a contains 16 groups, namely 71a, 111a, 102a, 83a, 174a, 145a,125a, 145b, 155a, 136a, 127a, 117b, 117c, 78a, 69a and 170a. In combination with Set 1, this modified set would give rise to 10 x 16 = 160 distinguishable components.

4. Extension of method to further dimensions.

The rules described so far allow two sets of substituents to be chosen which in combination give uniquely "self-coded" products. Set sizes of 10 are easily achieved, and diverse set sizes of up to 20 are not unrealistic. If a library contains a core structure with three variable sites, this can be accommodated by chosing the third set of substituents freely, and keeping the products as separate sublibraries, ie. the third variable is defined by its sublibrary. Since most assay systems involve some kind of partitioning (e.g. into wells) or spreading (e.g. over melanophore plates), there is little to be gained by mixing the sublibraries, compared to the structural information that would be lost in so doing. If Set 3 is of a similar size to Sets 1 and 2, it is clear that self-coded libraries in the range 1000 (10x10x10) to 8000 (20x20x20) components can be designed.

If a core structure contains four variable sites a further coding stage is required, ie. Sets 1, 2 and 3 coded, Set 4 as sublibraries. Selection of a third set of substituents may be made following Rule 7 below.

Rule 7. Exclude groups containing Cl and Br from Sets 1 and 2. Members of Set 3 should be selected freely provided that each member contains different numbers of Cl and Br atoms or other atoms distinguishable by their isotope patterns.

Example 4.1 Application of Rule 7 to Chart 4

The unmodified ten substituent sets, Set 1 and Set 2, selected by Rules 2 and 3 above, are suitable since neither contains chlorinated or brominated groups. A possible selection for Set 3 would comprise groups 69b (0 halogens), 111a (1 chlorine), 145b (2 chlorines), 117c (3 chlorines) and 155a (1 bromine). In

combination with an appropriate core structure, Sets 1, 2 and 3 would give a total of 10 x 10 x 5 = 500 components distinguishable by their nominal mass and isotopic mass patterns.

The application of Rule 7 forces Set 3 to contain mostly halogenated groups. However, further reagents containing artificially isotopically enriched substituents could be incorporated. For example, mixing CH_3COCl and CD_3COCl would give a substituent methyl group with nominal mass (M) 15 and isotopic mass at M+3.

It will be appreciated that in certain circumstances, especially in the case of smaller libraries, minor departures from any of the above rules may still provide libraries with zero or little mass redundancy.

5. Mathematical representation of the selection method

In additon to the graphical representation of the selection method using charts as described above, the selection rules, Rule 1 to Rule 3 may be represented mathematically.

The method requires that all numerical values used (molecular weight, periodicity, multiples and remainders) are integer values.

The periodicity, **P** corresponds to the number of columns in the graphical method.

The selection rules may now be expressed thus.

Rule 1. Set 1 and Set 2 are selected using the same value of **P** in Rule 2 and Rule 3 respectively.

Rule 2. Members of Set 1 are chosen such that all members have different molecular weights, and that these weights each give the **same** remainder (rem) when divided by **P**.

Rule 3. Members of Set 2 are chosen such that all members have different molecular weights, and that these weights each give a **different** remainder (rem) when divided by **P**.

Since the method described can be expressed in a mathematical form, the method can be incorporated into a computer program by means of an appropriate algorithm. Such a computerised implementation of the method falls within the scope of this

patent which in a further aspect provides a computer-based method for selection of reagents to construct a combinatorial library of compounds each having a unique molecular weight or isotope pattern, comprising the steps of:

a) selection, from databases of reagents, of those reagents suitable for the desired chemical transformation;

b) distinguishing substituent groups within each reagent structure and calculating the molecular weight and isotope pattern of each substituent group;

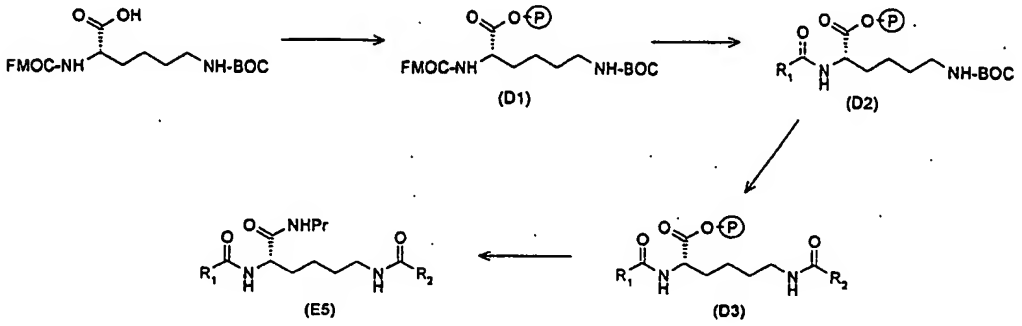
c) mapping the substituent groups according to their molecular weight into a chart format following the procedures described herein; and

d) allowing selection of reagents from the charts, either by manual selection or by an automated process according to the Rules described herein, and subsequent transfer of the reagents to an experimental worksheet for implementation of a manual or automated synthesis of the combinatorial library.

In a further aspect the invention provides a library synthesised using the above method, rules or algorithm and the use of the method, rules or algorithm for the synthesis of a chemical library.

The following examples illustrate the invention.

Example 5. Synthesis and analysis of a library with 140 distinguishable components (E5)



Description 1: Polymer-bound N(alpha)-FMOC-N(epsilon)-BOC-(*S*)-lysine (D1)

A solution of N(alpha)-FMOC-N(epsilon)-BOC-(*S*)-lysine (21.5 g, 44.8 mmol) in DCM (250 ml) was treated with 1,3-diisopropylcarbodiimide (5.65 g, 44.8 mmol) and DMAP (0.25 equivalents). After 30 min at room temperature, the solution was added to a suspension of 100-200 mesh hydroxymethyl polystyrene (10 g, 10 mmol) in DMF (50 ml). After agitating for 2.5 h the resin was washed well with DMF and DCM, then dried, to give the title compound, D1.

Description 2: Mixture of polymer-bound N(alpha)-acyl-N(epsilon)-BOC-(*S*)-lysines (D2)

In ten separate experiments, the polymer-bound lysine derivative from Description 1 (220 mg, 0.15 mmol) was washed with DCM (2x15 ml) and DMF (2x20 ml) then treated with 20% piperidine in DMF (2x20 ml) for 1 and 20 min. The resin was washed with DMF (2x15 ml) and DCM (2x15 ml), then treated for 4 hours with triethylamine (0.42 ml, 3 mmol) and one of the acid chlorides listed in Table 1 (1.5 mmol) in DCM (15 ml). The resin was washed with DMF (2x15 ml) and DCM (2x15 ml) then the portions were combined as a slurry in methanol and washed well with DCM to give the title compounds, D2 (2.14 g, approx 1.5 mmol) as a mixture of ten polymer-bound components.

Description 3: Mixtures of polymer-bound N(alpha)-acyl-N(epsilon)-acyl-(*S*)-lysines (D3)

In fourteen separate experiments, the polymer-bound lysine derivative from Description 2 (140 mg, approx. 0.1 mmol) was washed with DCM (2x15 ml), then treated with 30% trifluoroacetic acid and 2% anisole in DCM (2x15 ml) for 1 and 30 min. The resin was washed with DCM (3x15 ml), 10% triethylamine in DCM (2x10 ml), then was treated for 4 h with triethylamine (0.28 ml, 2 mmol) and one of the acid chlorides listed in Table 2 (1 mmol), in DCM (15 ml). The resin was washed with DMF (15 ml) and DCM (3x15 ml) and finally methanol (15 ml) to give the title compounds, D3, as fourteen mixtures each of ten polymer-bound components.

Analysis of libraries (E5) by Mass Spectroscopy

Individual beads were taken from libraries of Description 3 and treated with 50% n-propylamine in DCM (50 ul) for 6h. The reagents were allowed to evaporate, to give individual members of Example 5. Representative results of LC-MS analysis, along with the assignment of the molecular ions detected are given in Table 3.

TABLE 1

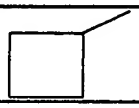
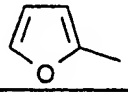
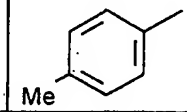
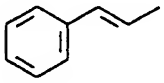
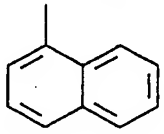
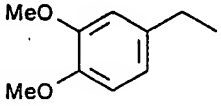
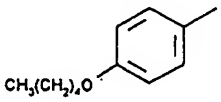
ENTRY	REAGENT (R_1-COCL)	STRUCTURE OF R_1	NOMINAL MASS R_1	MASS (R_1)/12
A	Butyryl chloride	$CH_3(CH_2)_2$-	43	3 rem 7
B	Cyclobutanecarbonyl chloride		55	4 rem 7
C	2-Furoyl chloride		67	5 rem 7
D	Toluoyl chloride		91	7 rem 7
E	Cinnamoyl chloride		103	8 rem 7
F	Methyl adipyl chloride	$MeO_2C(CH_2)_4$-	115	9 rem 7
G	1-Naphthoyl chloride		127	10 rem 7
H	10-Undecenoyl chloride	CH_2=CH$(CH_2)_8$-	139	11 rem 7
I	3,4-Dimethoxyphenyl-acetyl chloride		151	12 rem 7
J	4-Pentyloxybenzoyl chloride		163	13 rem 7

TABLE 2

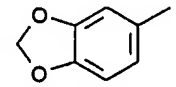
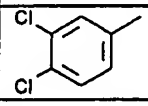
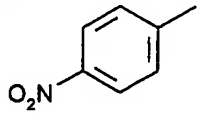
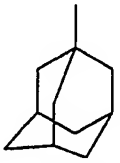
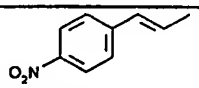
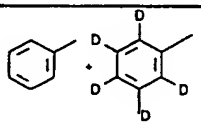
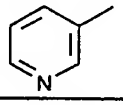
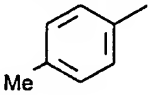
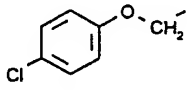
ENTRY	REAGENT (R_2-COCl)	STRUCTURE OF R_2	NOMINAL MASS OF R_2	MASS (R_2)/12
A	Piperonoyl chloride		121	10 rem 1
B	3,4-dichloro-benzoyl chloride		145	12 rem 1
C	4-Nitrobenzoyl chloride		122	10 rem 2
D	1-Adamantane-carbonyl chloride		135	11 rem 3
E	Trans-4-nitro-cinnamoyl chloride		148	12 rem 4
F	Propionyl chloride	CH_3CH_2-	29	2 rem 5
G	1:1 Benzoyl chloride/Penta-deuterobenzoyl chloride		77	6 rem 5
H	Nicotinoyl chloride		78	6 rem 6
I	4-Toluoyl chloride		91	7 rem 7
J	4-Chloro-phenoxyacetyl chloride		141	11 rem 9

TABLE 2 cont'd

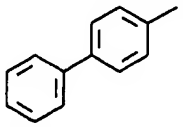
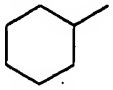
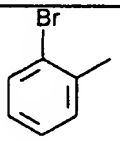
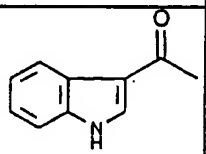
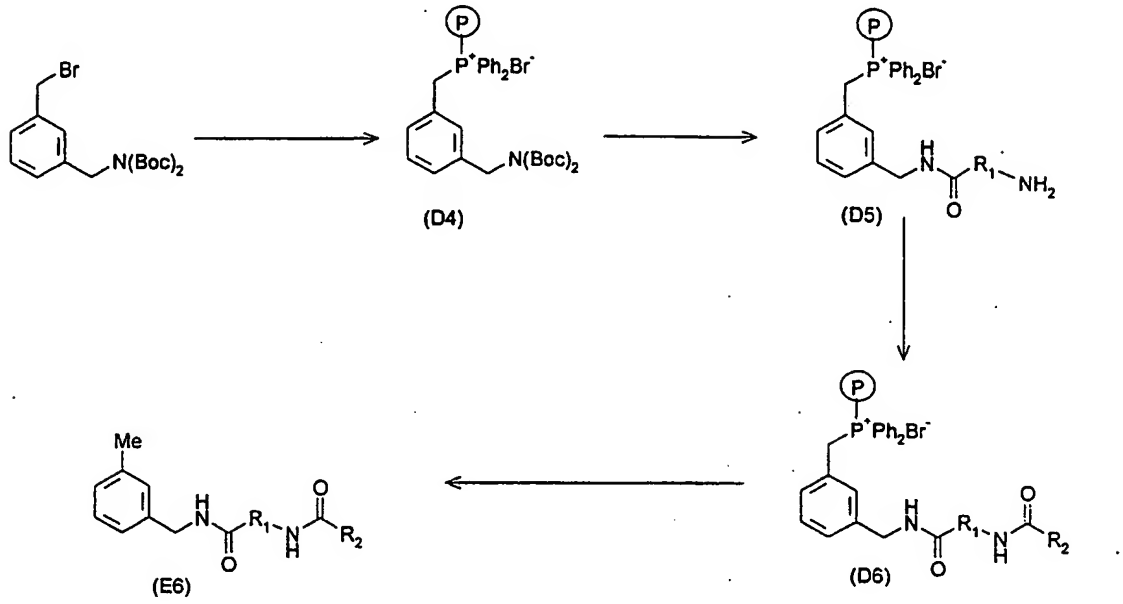
K	4-Biphenyl-carbonyl chloride		153	12 rem 9
N	Cyclohexane-carbonyl chloride		83	6 rem 11
M	2-Bromobenzoyl chloride		155	12 rem 11
N	Indol-3-glyoxylyl chloride		144	12 rem 0

TABLE 3

	Assignment		
M+H (found)	R_1 (TABLE 1)	R_2 (TABLE 2)	MASS (calc)
527	J	C	526
428	E	L	427
488	J	L	487
362	D	F	361
548/550 (1:1)	I	M	547/549(1:1)
446/451(1:1)	G	G	445/450(1:1)

Example 6. Synthesis and analysis of a library with 50 distinguishable components (E6)



Description 4: Polymer-bound 3-[N,N-di(t-butyloxycarbonyl)-aminomethyl]benzyl triphenylphosphonium bromide (D4)

A suspension of polymer-bound triphenylphosphine prepared from 150-200 um 4-bromopolystyrene by the method of Bernard and Ford (J. Org. Chem. **1983**, *48*, 326) (2.96 g, approx 5 meq) in DMF (25 ml) was treated with N,N-di(t-butyloxycarbonyl)-3-bromomethylbenzylamine (4 g, 10 mmol). The mixture was heated at 70 °C for 48 h, cooled, filtered and the polymer was washed alternately with toluene (50 ml) and DCM (50 ml) (3 times), then DCM (50 ml) and ether (50 ml) (3 times), and finally with 1:1 DCM/ether (50 ml) and ether (2x50 ml) to give the title compound, D4 (4.21 g).

Description 5: Mixture of polymer-bound 3-[N-(aminoacyl)aminomethyl]benzyl triphenylphosphonium salts (D5)

In five separate experiments, the polymer-bound phosphonium salt from Description 4 (385 mg, 0.35 meq) was washed with DCM (2x25 ml), then treated with a mixture of 30% trifluoroacetic acid and 2% anisole in DCM (2x25 ml) for 1 and 30 min. The resin was washed with DCM (3x25 ml), 10% triethylamine in DCM (2x25 ml) and DCM (2x25 ml). The resin was suspended overnight in a solution of 1-hydroxy-7-azabenzotriazole (109 mg, 0.8 mmol), 1,3-

diisopropylcarbodiimide (0.13 ml, 0.8 mmol) and one of the carboxylic acids listed in Table 4 (0.75 mmol) in DMF (10 ml) and DCM (3 ml). After washing with DMF (2x25 ml) and DCM (3x25 ml), the five products were combined as a slurry in DCM, filtered and treated with 20% piperidine in DMF (2x30 ml) for 1 and 30 min. The resin was washed with DMF (2x30 ml) and DCM (3x30 ml), then dried to give the title compounds, D5 (1.83 g, approx 1.75 mmol) as a mixture of five polymer-bound components..

Description 6: Mixture of polymer-bound 3-[N-(acylaminoacyl)aminomethyl]benzyl triphenylphosphonium salts (D6)

In ten separate experiments, the polymer-bound phosphonium salt from Description 5 (160 mg, approx 0.17 mmol) was suspended in DCM (10 ml) and treated with triethylamine (0.1 ml, 0.7 mmol) and one of the acid chlorides listed in Table 5 (0.35 mmol). After 90 min the resin was filtered and washed with DCM (2x20 ml). The ten products were combined as a slurry in DCM and washed with DMF (3x30 ml) and DCM (2x30 ml). The resin was finally washed successively with 3:1, 1:1 and 1:3 mixtures of DCM and ether, and then with ether alone to give the title compounds, D6, as a mixture of 50 polymer-bound components.

Analysis of library (E6) by mass spectroscopy.

Individual beads from Description 6 were treated with a 10 mM solution of sodium hydroxide in 10% aqueous dioxane (50 ul) overnight to effect cleavage to the individual members of Example 6. Representative results of APCI-MS analysis, along with the assignment of the molecular ions detected are given in Table 6.

TABLE 4

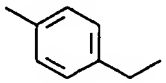
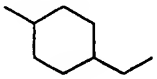
ENTRY	REAGENT HO_2C-R_1-NHFMOC	STRUCTURE OF R_1	NOMINAL MASS R_1	MASS $(R_1)/12$
A	N-FMOC-glycine	$-CH_2-$	14	1 rem 2
B	N-FMOC-beta-alanine	$-(CH_2)_2-$	28	2 rem 4
C	N-FMOC-6-amino hexanoic acid	$-(CH_2)_5-$	70	5 rem 10
D	N-FMOC-4-aminomethyl benzoic acid		90	7 rem 6
E	N-FMOC-4-aminomethyl cyclohexanecarboxylic acid		96	8 rem 0

TABLE 5

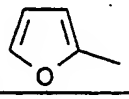
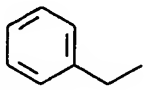
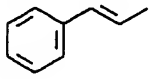
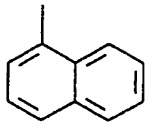
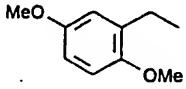
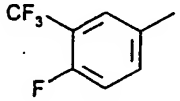
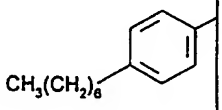
ENTRY	REAGENT (R_2-COCl)	STRUCTURE OF R_2	NOMINAL MASS OF R_2	MASS (R_2)/12
A	Butryl chloride	$CH_3(CH_2)_2$-	43	3 rem 7
B	2-Furoyl chlorde		67	5 rem 7
C	Phenylacetyl chloride		91	7 rem 7
D	Cinnamoyl chloride		103	8 rem 7
E	Methyl adipyl chloride	$MeO_2C(CH_2)_4$-	115	9 rem 7
F	1-naphthoyl chloride		127	10 rem 7
G	10-undecenoyl chloride	CH_2=CH-$(CH_2)_8$-	139	11 rem 7
H	2,5-dimethoxy-phenylacetyl chloride		151	12 rem 7
I	4-Fluoro-3-(trifluoromethyl)-benzoyl chloride		163	13 rem 7
J	4-n-heptylbenzoyl chloride		175	14 rem 7

TABLE 6

	ASSIGNMENT		
M+H (found)	R1 (TABLE 3)	R2 (TABLE 4)	MASS (calc)
365	C	D	364
329	C	B	328
379	E	C	378
439	E	H	438
391	E	D	390
409	D	F	408
329	C	B	328
457	D	J	456
325	D	A	324
357	A	H	356
403	E	E	402
397	D	E	396
433	D	H	432
359	B	G	358
353	C	C	352
325	D	A	324
403	E	E	402
427	E	G	426

CLAIMS:

1. A method for the control of mass redundancies in a combinatorially synthesised compound library which comprises identifying compounds by their molecular weight.

2. The method according to claim 1 where molecular weight is determined by mass spectrometry.

3. A library of compounds designed and synthesised using the method of claim 1 or 2.

4. Use of the method according to claim 1 or 2 for the design and synthesis of a chemical library.

5. A computer-based method for selection of reagents to construct a combinatorial library of compounds each having a unique molecular weight or isotope pattern, comprising the steps of:

a) selection, from databases of reagents, of those reagents suitable for the desired chemical transformation;

b) distinguishing substituent groups within each reagent structure and calculating the molecular weight and isotope pattern of each substituent group;

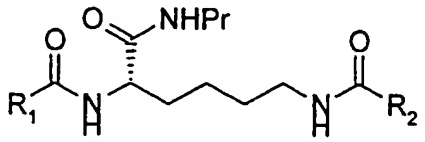
c) mapping the substituent groups according to their molecular weight into a chart format following the procedures described herein; and

d) allowing selection of reagents from the charts, either by manual selection or by an automated process according to the Rules defined herein, and subsequent transfer of the reagents to an experimental worksheet for implementation of a manual or automated synthesis of the combinatorial library

6. A library of compounds designed and synthesised using the computer-based method of claim 5.

7. Use of the computer-based method according to claim 5 for the design and synthesis of a chemical library.

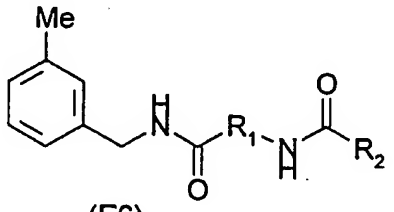
6. A library comprising two or more compounds of formula (E5):



(E5)

where R_1 and R_2 are as defined in example 5.

7. A library comprising two or more compounds of formula (E6):



(E6)

where R_1 and R_2 are as defined in example 6

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 1/04, C07C 231/00 — A3

(11) International Publication Number: WO 97/08190

(43) International Publication Date: 6 March 1997 (06.03.97)

(21) International Application Number: PCT/EP96/03731

(22) International Filing Date: 23 August 1996 (23.08.96)

(30) Priority Data:
9517661.6 30 August 1995 (30.08.95) GB

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: HUGHES, Ian [GB/GB]; SmithKline Beecham Pharmaceuticals, New Frontiers Science Park South, Third Avenue, Harlow, Essex CM19 5AW (GB).

(74) Agent: SUMMERSELL, Richard, John; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 27 March 1997 (27.03.97)

(54) Title: COMPOUNDS

(57) Abstract

The invention relates to combinatorial chemistry, in particular synthesis of combinatorial libraries.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

INTERNATIONAL SEARCH REPORT

International Application No
PL./EP 96/03731

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C07K1/04 C07C231/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07C C07D C07H C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 95 19359 A (MASSACHUSETTS INST TECHNOLOGY) 20 July 1995 see page 8; claims	1,3,6,7
A	WO 95 04160 A (ISIS INNOVATION ;SOUTHERN EDWIN (GB); CUMMINS WILLIAM JONATHAN (GB)) 9 February 1995 see page 49; claims	1,3,6,7
A	US 5 288 514 A (ELLMAN JONATHAN A) 22 February 1994 see abstract; claims	1,3,6,7
A	WO 94 08051 A (UNIV COLUMBIA ;COLD SPRING HARBOR LAB (US); STILL W CLARK (US); OH) 14 April 1994 see claims	1,3,6,7

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
10 February 1997	21 -02- 1997

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Day, G

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 96/03731

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9519359	20-07-95	AU-A-	1562895	01-08-95
WO-A-9504160	09-02-95	AU-A-	7269194	28-02-95
		CA-A-	2168010	09-02-95
		CN-A-	1131440	18-09-96
		EP-A-	0711362	15-05-96
		FI-A-	960403	29-01-96
		HU-A-	73802	30-09-96
		NO-A-	960370	28-03-96
US-A-5288514	22-02-94	AU-A-	4923193	12-04-94
		WO-A-	9406291	31-03-94
		US-A-	5545568	13-08-96
WO-A-9408051	14-04-94	AU-A-	5536994	26-04-94
		CA-A-	2143848	14-04-94
		EP-A-	0665897	09-08-95
		HU-A-	72495	28-05-96
		JP-T-	8506175	02-07-96
		NO-A-	951230	30-03-95
		US-A-	5565324	15-10-96